

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



THỰC TẬP CƠ SỞ

BÁO CÁO GIỮA KỲ

Đề tài: XÂY DỰNG HỆ THỐNG GỢI Ý SẢN PHẨM

Giảng viên	: Kim Ngọc Bách
Họ và tên sinh viên	: Nguyễn Văn Trường
Mã sinh viên	: B23DCCN870
Lớp	: D23CQCN02-B
Nhóm	: 11

Hà Nội – 2026

Mục lục

Nội dung

Mục lục	1
Danh mục thuật ngữ, viết tắt	2
1. Giới thiệu	3
1.1 Lý do chọn đề tài	3
1.2 Ý nghĩa của đề tài	3
1.3 Tính ứng dụng	5
1.4 Giá trị học thuật	5
2. Cơ sở lý thuyết và công nghệ sử dụng	6
2.1 Tổng quan về hệ thống gợi ý (Recommender Systems)	6
2.2 Collaborative Filtering	7
2.3 Matrix Factorization	8
3. Phân tích yêu cầu của dự án	9
3.1 Mục tiêu hệ thống	9
3.2 Yêu cầu chức năng (Functional Requirements)	9
3.3 Yêu cầu phi chức năng (Non-functional Requirements)	10
3.4 Phạm vi dự án	11
3.5 Các thách thức dự kiến	11
4. Kế hoạch thực hiện dự án	11
4.1 Tổng quan kế hoạch	11
4.2 Kế hoạch chi tiết theo từng giai đoạn	12
Giai đoạn 1: Nghiên cứu lý thuyết	12
Giai đoạn 2: Thu thập và tiền xử lý dữ liệu	12
Giai đoạn 3: Xây dựng mô hình	13
Giai đoạn 4: Đánh giá và tối ưu	13
Giai đoạn 5: Xây dựng demo hệ thống	13
Giai đoạn 6: Hoàn thiện báo cáo	14
4.3 Công cụ quản lý và hỗ trợ	14
4.4 Rủi ro và phương án xử lý	14
5. Danh mục tài liệu tham khảo	15

Danh mục thuật ngữ, viết tắt

Thuật ngữ	Giải thích
Recommender System(RS)	Hệ thống thông minh dự đoán và đề xuất sản phẩm/nội dung phù hợp với người dùng
Information Overload	Hiện tượng quá tải thông tin khi người dùng phải lựa chọn giữa quá nhiều phương án
Collaborative Filtering(CF)	Phương pháp gợi ý dựa trên hành vi tương tự giữa người dùng hoặc sản phẩm
Content-based Filtering	Phương pháp gợi ý dựa trên đặc trưng nội dung của sản phẩm
Hybrid Recommendation	Phương pháp kết hợp nhiều kỹ thuật gợi ý
Matrix Factorization(MF)	Kỹ thuật phân rã ma trận user-item thành các đặc trưng tiềm ẩn
Latent Factors	Các đặc trưng tiềm ẩn biểu diễn sở thích người dùng hoặc thuộc tính sản phẩm
User-Item Matrix	Ma trận biểu diễn tương tác giữa người dùng và sản phẩm
Data Sparsity	Hiện tượng ma trận user-item có rất nhiều giá trị trống
Cold-start Problem	Vấn đề người dùng/sản phẩm mới chưa có dữ liệu
Overfitting	Hiện tượng mô hình học quá sát dữ liệu huấn luyện
Top-N Recommendation	Bài toán đề xuất N sản phẩm phù hợp nhất cho người dùng
Rating Prediction	Bài toán dự đoán điểm đánh giá
Loss Function	Hàm mất mát dùng để tối ưu mô hình
Regularization	Kỹ thuật điều chuẩn nhằm tránh overfitting
Gradient Descent	Thuật toán tối ưu dựa trên đạo hàm
Alternating Least Squares (ALS)	Thuật toán tối ưu thay phiên hai ma trận trong MF
Neural Collaborative Filtering	Mô hình CF kết hợp mạng nơ-ron
Deep Learning-based Recommendation	Hệ thống gợi ý sử dụng học sâu

Thuật ngữ	Giải thích
Conversion Rate	Tỷ lệ chuyển đổi từ người xem sang mua hàng
Exploratory Data Analysis (EDA)	Phân tích dữ liệu khám phá
Root Mean Squared Error(RMSE)	Sai số căn bậc hai trung bình
Machine Learning(ML)	Học máy

1. Giới thiệu

1.1 Lý do chọn đề tài

Trong bối cảnh chuyển đổi số diễn ra mạnh mẽ, sự gia tăng nhanh chóng về số lượng sản phẩm và nội dung trên các nền tảng trực tuyến đã đặt ra thách thức lớn đối với người dùng. Việc phải lựa chọn giữa hàng triệu sản phẩm khác nhau dẫn đến hiện tượng **quá tải thông tin (Information Overload)**, làm gia tăng chi phí tìm kiếm, gây khó khăn trong quá trình ra quyết định và ảnh hưởng tiêu cực đến trải nghiệm khách hàng.

Để giải quyết vấn đề này, nhiều doanh nghiệp công nghệ hàng đầu đã ứng dụng **Hệ thống gợi ý (Recommender Systems)** như một giải pháp trọng tâm trong chiến lược phát triển nền tảng số. Các hệ thống này sử dụng dữ liệu hành vi và lịch sử tương tác của người dùng để phân tích sở thích cá nhân, từ đó đưa ra các đề xuất phù hợp. Thực tế cho thấy, mô hình này đã đóng vai trò quan trọng trong sự thành công của các nền tảng lớn như Amazon, Netflix và Shopee, khi góp phần nâng cao mức độ hài lòng của người dùng, tăng tỷ lệ chuyển đổi và tối ưu hóa doanh thu.

Nhận thấy tầm quan trọng cả về mặt lý luận và thực tiễn của bài toán cá nhân hóa nội dung, đề tài “**Xây dựng hệ thống gợi ý sản phẩm sử dụng Collaborative Filtering và Matrix Factorization**” được lựa chọn nghiên cứu. Đề tài tập trung khai thác dữ liệu tương tác giữa người dùng và sản phẩm nhằm xây dựng mô hình dự đoán sở thích cá nhân một cách chính xác. Thông qua việc ứng dụng các kỹ thuật phân rã ma trận hiện đại, hệ thống hướng đến việc nâng cao chất lượng gợi ý, đồng thời góp phần làm rõ các nguyên lý cốt lõi trong lĩnh vực Học máy và Hệ thống gợi ý.

1.2 Ý nghĩa của đề tài

Về mặt thực tiễn, hệ thống gợi ý đóng vai trò quan trọng trong việc giải quyết vấn đề quá tải thông tin trên các nền tảng thương mại điện tử và dịch vụ trực tuyến. Trong bối cảnh số lượng sản phẩm và nội dung ngày càng gia tăng, người dùng gặp khó khăn trong việc lựa chọn sản phẩm phù hợp. Hệ thống gợi ý cho phép tự động phân tích hành vi, lịch sử tương tác và sở thích cá nhân để đưa ra các đề xuất phù hợp, thay vì yêu cầu người dùng tự tìm kiếm trong một danh sách lớn các lựa chọn. Nhờ đó, hệ thống góp phần:

- Nâng cao mức độ hài lòng của người dùng
- Cải thiện trải nghiệm cá nhân hóa
- Gia tăng tỷ lệ chuyển đổi và tối ưu doanh thu cho doanh nghiệp

Thực tế cho thấy các nền tảng như Amazon, Netflix và Shopee đã ứng dụng thành công hệ thống gợi ý như một thành phần cốt lõi trong chiến lược phát triển và duy trì người dùng.

Về mặt công nghệ và học thuật, đề tài tạo điều kiện nghiên cứu chuyên sâu các phương pháp học máy phổ biến trong lĩnh vực Recommendation Systems, đặc biệt là Collaborative Filtering và Matrix Factorization. Thông qua quá trình xây dựng, huấn luyện và đánh giá mô hình, đề tài giúp:

- Hiểu rõ cách biểu diễn và xử lý dữ liệu tương tác người dùng – sản phẩm (user-item matrix)
- Phân tích và khai thác các đặc trưng tiềm ẩn (latent factors) trong dữ liệu thưa
- Tối ưu hóa mô hình bằng các thuật toán như Gradient Descent hoặc Alternating Least Squares (ALS)
- Đánh giá hiệu quả hệ thống thông qua các chỉ số định lượng như RMSE, Precision@K và Recall@K

Nhờ đó, đề tài không chỉ có giá trị ứng dụng thực tế mà còn góp phần xây dựng nền tảng kiến thức vững chắc cho việc nghiên cứu và phát triển các mô hình gợi ý nâng cao trong lĩnh vực Trí tuệ nhân tạo và Học máy.

1.3 Tính ứng dụng

Hệ thống gợi ý sản phẩm là một trong những ứng dụng tiêu biểu và có tính thực tiễn cao của Machine Learning trong kỷ nguyên dữ liệu lớn. Nhiều nghiên cứu đã chỉ ra rằng hệ thống gợi ý đóng vai trò quan trọng trong việc cá nhân hóa nội dung, hỗ trợ ra quyết định và tối ưu hóa trải nghiệm người dùng trên các nền tảng số [1]. Do đó, việc xây dựng hệ thống gợi ý không chỉ mang ý nghĩa nghiên cứu mà còn có khả năng triển khai trực tiếp trong các hệ thống thực tế.

Trong lĩnh vực thương mại điện tử, các nền tảng lớn như Amazon và Shopee ứng dụng hệ thống gợi ý để đề xuất sản phẩm dựa trên lịch sử mua hàng, lượt xem và đánh giá của người dùng. Theo các báo cáo kỹ thuật và nghiên cứu thực nghiệm, hệ thống gợi ý có thể đóng góp đáng kể vào doanh thu thông qua việc gia tăng tỷ lệ chuyển đổi (conversion rate) và giá trị đơn hàng trung bình [2].

Trong lĩnh vực giải trí số, các nền tảng như Netflix và YouTube sử dụng hệ thống đề xuất để gợi ý phim, chương trình hoặc video phù hợp với sở thích cá nhân. Việc cá nhân hóa nội dung này góp phần gia tăng thời gian sử dụng dịch vụ, cải thiện mức độ tương tác và nâng cao khả năng giữ chân người dùng trong dài hạn [3].

Bên cạnh đó, hệ thống gợi ý còn được ứng dụng rộng rãi trong nhiều lĩnh vực khác:

- **Giáo dục trực tuyến:** Gợi ý khóa học, tài liệu hoặc lộ trình học tập phù hợp với năng lực và mục tiêu của người học [1].
- **Mạng xã hội:** Đề xuất bạn bè, nhóm quan tâm hoặc nội dung cá nhân hóa dựa trên hành vi tương tác.
- **Marketing thông minh:** Cá nhân hóa quảng cáo và chiến dịch tiếp thị dựa trên phân tích hành vi và sở thích khách hàng.

Ngoài phạm vi nghiên cứu, đề tài có thể được mở rộng và tích hợp vào các hệ thống thực tế như website bán hàng, ứng dụng di động hoặc nền tảng phân tích dữ liệu doanh nghiệp. Nhờ tính linh hoạt, khả năng mở rộng và hiệu quả đã được kiểm chứng trong nhiều lĩnh vực, hệ thống gợi ý sản phẩm có tiềm năng ứng dụng rộng rãi trong đời sống và hoạt động kinh doanh hiện đại.

1.4 Giá trị học thuật

Đề tài “**Xây dựng hệ thống gợi ý sản phẩm sử dụng Collaborative Filtering và Matrix Factorization**” mang giá trị học thuật quan trọng trong lĩnh vực Machine Learning và Khai phá dữ liệu (Data Mining).

Thứ nhất, đề tài tiếp cận một trong những bài toán cốt lõi của lĩnh vực Recommendation Systems – bài toán dự đoán sở thích người dùng dựa trên dữ liệu tương tác giữa người dùng và sản phẩm (user–item interaction). Đây là nền tảng của nhiều hệ thống thông minh hiện đại và đã được nghiên cứu rộng rãi trong cộng đồng khoa học [1].

Thứ hai, việc triển khai các phương pháp như Collaborative Filtering và Matrix Factorization giúp làm rõ các khái niệm học thuật quan trọng trong hệ thống gợi ý. Đặc biệt, Matrix Factorization được xem là một trong những kỹ thuật hiệu quả và có ảnh hưởng sâu rộng trong bài toán dự đoán rating và cá nhân hóa nội dung [4]. Thông qua quá trình xây dựng mô hình, đề tài làm sáng tỏ các vấn đề kỹ thuật sau:

- Biểu diễn dữ liệu dưới dạng ma trận người dùng – sản phẩm (user–item matrix)
- Đặc tính thưa của dữ liệu (data sparsity) trong môi trường thực tế
- Phân rã ma trận thành các đặc trưng tiềm ẩn (latent factors) để khám phá sở thích ẩn
- Tối ưu hóa hàm mất mát (loss function) bằng các thuật toán như Gradient Descent hoặc Alternating Least Squares
- Đánh giá hiệu năng mô hình bằng các chỉ số định lượng như RMSE, Precision@K và Recall@K[6]

Thứ ba, đề tài tạo nền tảng cho việc mở rộng nghiên cứu sang các hướng tiếp cận hiện đại hơn như Neural Collaborative Filtering và các mô hình gợi ý dựa trên Deep Learning, vốn đang là xu hướng phát triển mạnh mẽ trong những năm gần đây [5].

2. Cơ sở lý thuyết và công nghệ sử dụng

2.1 Tổng quan về hệ thống gợi ý (Recommender Systems)

Hệ thống gợi ý (Recommender Systems) là các hệ thống thông minh có nhiệm vụ dự đoán mức độ quan tâm của người dùng đối với một sản phẩm hoặc nội dung dựa trên dữ liệu lịch sử tương tác. Theo Ricci et al. [1], hệ thống gợi ý đóng vai trò quan trọng trong việc hỗ trợ người dùng ra quyết định trong môi trường thông tin lớn và phức tạp.

Hệ thống gợi ý thường được phân thành ba nhóm chính [1]:

- **Content-based Filtering**
- **Collaborative Filtering**
- **Hybrid Methods**

Trong phạm vi đề tài này, phương pháp **Collaborative Filtering** và **Matrix Factorization** được lựa chọn làm nền tảng xây dựng mô hình do tính hiệu quả và khả năng mở rộng cao trong thực tế.

2.2 Collaborative Filtering

Collaborative Filtering (CF) là phương pháp dựa trên giả định rằng những người dùng có hành vi tương tự trong quá khứ sẽ có xu hướng yêu thích các sản phẩm tương tự trong tương lai [1].

CF thường được chia thành hai hướng chính:

- **User-based Collaborative Filtering**
- **Item-based Collaborative Filtering**

Phương pháp Item-based Collaborative Filtering được ứng dụng hiệu quả trong hệ thống gợi ý của Amazon [2].

CF sử dụng các độ đo tương đồng phổ biến như:

- Cosine Similarity
- Pearson Correlation

Ưu điểm của CF là không yêu cầu thông tin nội dung sản phẩm mà chỉ cần dữ liệu tương tác người dùng. Tuy nhiên, phương pháp này gặp phải các vấn đề như:

- Dữ liệu thưa (data sparsity)
- Cold-start (người dùng/sản phẩm mới)

2.3 Matrix Factorization

Matrix Factorization (MF) là một trong những phương pháp hiệu quả nhất trong hệ thống gợi ý hiện đại. Phương pháp này phân rã ma trận người dùng – sản phẩm thành hai ma trận đặc trưng tiềm ẩn (latent factor matrices).

Koren et al. [4] đã chứng minh rằng Matrix Factorization đạt hiệu suất vượt trội trong cuộc thi Netflix Prize nhờ khả năng mô hình hóa sở thích người dùng và đặc điểm sản phẩm dưới dạng vector đặc trưng tiềm ẩn. Thành công của hệ thống gợi ý tại Netflix cũng được phân tích chi tiết trong [3].

Giả sử R là ma trận đánh giá, ta thực hiện phân rã:

$$R \approx P \times Q^T$$

Trong đó:

P : Ma trận đặc trưng người dùng

Q : Ma trận đặc trưng sản phẩm

Việc tối ưu thường được thực hiện bằng cách tối thiểu hóa hàm mất mát thông qua Gradient Descent hoặc Alternating Least Squares [4].

2.4 Các chỉ số đánh giá mô hình

Để đánh giá hiệu quả hệ thống gợi ý, các chỉ số phổ biến bao gồm:

RMSE (Root Mean Squared Error) – dùng cho bài toán dự đoán rating

Precision@K – đo chất lượng danh sách gợi ý

Recall@K – đo khả năng bao phủ các mục liên quan

Theo Herlocker et al. [6], việc lựa chọn chỉ số đánh giá phù hợp phụ thuộc vào mục tiêu của hệ thống (rating prediction hay top-N recommendation).

2.5 Công nghệ sử dụng

Hệ thống được triển khai bằng các công nghệ sau:

Ngôn ngữ lập trình: **Python**

Thư viện xử lý dữ liệu: **NumPy**, **Pandas**

Thư viện Machine Learning: **Scikit-learn**

Thư viện xây dựng hệ thống gợi ý: **Surprise** (tùy chọn)

Triển khai hệ thống: **Flask** hoặc **FastAPI**

Python được lựa chọn nhờ tính linh hoạt, hệ sinh thái phong phú và khả năng hỗ trợ mạnh mẽ cho các bài toán Machine Learning [7].

3. Phân tích yêu cầu của dự án

3.1 Mục tiêu hệ thống

Hệ thống được xây dựng nhằm:

- Dự đoán mức độ quan tâm của người dùng đối với sản phẩm
- Đề xuất danh sách sản phẩm phù hợp với từng người dùng (Top-N Recommendation)
- Tối ưu hóa độ chính xác của gợi ý dựa trên dữ liệu lịch sử

Mục tiêu cuối cùng là xây dựng một mô hình gợi ý có khả năng cá nhân hóa dựa trên phương pháp Collaborative Filtering và Matrix Factorization.

3.2 Yêu cầu chức năng (Functional Requirements)

Hệ thống cần đáp ứng các chức năng sau:

1. Thu thập và xử lý dữ liệu

- Đọc dữ liệu user-item (rating hoặc lịch sử tương tác)
- Làm sạch dữ liệu (loại bỏ giá trị thiếu, trùng lặp)

- Chuyển đổi dữ liệu về dạng ma trận user–item

2. Xây dựng mô hình gợi ý

- Tính toán độ tương đồng giữa người dùng hoặc sản phẩm (Collaborative Filtering)
- Triển khai Matrix Factorization để phân rã ma trận
- Huấn luyện mô hình bằng thuật toán tối ưu (ví dụ: Gradient Descent)

3. Dự đoán và đề xuất

- Dự đoán rating cho các sản phẩm chưa được đánh giá
- Đề xuất danh sách Top-N sản phẩm cho từng người dùng

4. Đánh giá mô hình

- Tính RMSE đối với bài toán dự đoán rating
- Tính Precision@K và Recall@K cho bài toán gợi ý Top-N

3.3 Yêu cầu phi chức năng (Non-functional Requirements)

- **Hiệu suất:** Mô hình phải xử lý được tập dữ liệu có kích thước trung bình (ví dụ: hàng nghìn người dùng và sản phẩm).
- **Khả năng mở rộng:** Có thể nâng cấp sang mô hình nâng cao (Deep Learning-based Recommendation).
- **Tính chính xác:** Kết quả gợi ý phải đạt độ chính xác ở mức chấp nhận được theo các chỉ số đánh giá.
- **Tính mô-đun:** Hệ thống được thiết kế tách biệt giữa xử lý dữ liệu, huấn luyện mô hình và đánh giá.

3.4 Phạm vi dự án

Trong phạm vi đề tài:

- Sử dụng dataset có sẵn (ví dụ: MovieLens hoặc dữ liệu mô phỏng)
- Tập trung vào Collaborative Filtering và Matrix Factorization
- Triển khai bằng Python và các thư viện ML

Ngoài phạm vi đề tài:

- Không xử lý dữ liệu thời gian thực (real-time streaming)
- Không triển khai hệ thống phân tán quy mô lớn
- Không xây dựng hệ thống thương mại hoàn chỉnh

3.5 Các thách thức dự kiến

- **Cold-start problem:** Người dùng mới chưa có dữ liệu
- **Data sparsity:** Ma trận user–item rất thưa
- **Overfitting:** Mô hình học quá sát dữ liệu huấn luyện

Việc nhận diện các thách thức này giúp định hướng giải pháp và cải tiến mô hình trong các giai đoạn tiếp theo.

4. Kế hoạch thực hiện dự án

Phần này trình bày lộ trình triển khai dự án theo từng giai đoạn, đảm bảo hoàn thành đúng tiến độ và đạt mục tiêu đề ra.

4.1 Tổng quan kế hoạch

Dự án được triển khai theo 6 giai đoạn chính:

1. Nghiên cứu lý thuyết
2. Thu thập và xử lý dữ liệu
3. Xây dựng mô hình
4. Đánh giá và tối ưu
5. Xây dựng demo (nếu có)
6. Viết báo cáo và hoàn thiện

4.2 Kế hoạch chi tiết theo từng giai đoạn

Giai đoạn 1: Nghiên cứu lý thuyết

- Tìm hiểu tổng quan về Recommender Systems
- Nghiên cứu Collaborative Filtering
- Nghiên cứu Matrix Factorization
- Tìm hiểu các chỉ số đánh giá: RMSE, Precision@K, Recall@K

Kết quả đạt được:

Hiểu rõ nền tảng lý thuyết và lựa chọn phương pháp phù hợp.

Giai đoạn 2: Thu thập và tiền xử lý dữ liệu

- Lựa chọn dataset
- Làm sạch dữ liệu
- Phân tích dữ liệu (EDA)

- Xây dựng ma trận user–item

Kết quả đạt được:

Dataset sẵn sàng cho việc huấn luyện mô hình.

Giai đoạn 3: Xây dựng mô hình

- Triển khai User-based hoặc Item-based Collaborative Filtering
- Triển khai Matrix Factorization
- Huấn luyện mô hình bằng Gradient Descent
- Chia tập train/test

Kết quả đạt được:

Mô hình có khả năng dự đoán rating và gợi ý Top-N.

Giai đoạn 4: Đánh giá và tối ưu

- Tính RMSE trên tập test
- Tính Precision@K và Recall@K
- Điều chỉnh siêu tham số (số latent factors, learning rate, regularization)

Kết quả đạt được:

Mô hình đạt hiệu suất tối ưu trong phạm vi đề tài.

Giai đoạn 5: Xây dựng demo hệ thống

- Xây dựng giao diện đơn giản (console hoặc web Flask/FastAPI)
- Cho phép nhập user ID và hiển thị danh sách sản phẩm gợi ý
- Kiểm tra hoạt động toàn hệ thống

Kết quả đạt được:

Hệ thống gợi ý có thể chạy thử nghiệm.

Giai đoạn 6: Hoàn thiện báo cáo

- Viết báo cáo chi tiết
- Trình bày kết quả thực nghiệm
- Phân tích ưu điểm và hạn chế
- Đề xuất hướng phát triển

Kết quả đạt được:

Hoàn thành báo cáo và chuẩn bị bảo vệ.

4.3 Công cụ quản lý và hỗ trợ

- Quản lý mã nguồn bằng Git
- Lập trình bằng Python
- Phân tích dữ liệu bằng Pandas, NumPy
- Xây dựng mô hình bằng Scikit-learn hoặc Surprise

4.4 Rủi ro và phương án xử lý

Rủi ro	Giải pháp
Mô hình không đạt độ chính xác mong muốn	Điều chỉnh siêu tham số, thử phương pháp khác
Dữ liệu quá thừa	Giảm kích thước dataset hoặc thử regularization

Thiếu thời gian	Ưu tiên hoàn thành mô hình cơ bản trước
-----------------	---

5. Danh mục tài liệu tham khảo

- [1] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 2nd ed. Springer, 2015.
- [2] G. Linden, B. Smith, and J. York, “Amazon.com recommendations: Item-to-item collaborative filtering,” *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.
- [3] C. A. Gomez-Urbe and N. Hunt, “The Netflix Recommender System: Algorithms, Business Value, and Innovation,” *ACM Transactions on Management Information Systems*, vol. 6, no. 4, 2015.
- [4] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [5] X. He et al., “Neural collaborative filtering,” in *Proceedings of the 26th International World Wide Web Conference (WWW 2017)*, 2017, pp. 173–182.
- [6] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5–53, 2004.
- [7] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.